

## 対応なし、英抄

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2002-73680

(P2002-73680A)

(43) 公開日 平成14年3月12日 (2002.3.12)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テーマコード* (参考)
G 0 6 F 17/30	3 5 0	G 0 6 F 17/30	3 5 0 C 5 B 0 7 5
	1 7 0		1 7 0 A
	3 3 0		3 3 0 C

審査請求 有 請求項の数 8 O L (全 21 頁)

(21) 出願番号 特願2000-261366(P2000-261366)

(22) 出願日 平成12年8月30日 (2000.8.30)

(71) 出願人 591115475

株式会社三菱総合研究所

東京都千代田区大手町2丁目3番6号

(72) 発明者 中 村 達 生

東京都千代田区大手町2-3-6 株式会  
社三菱総合研究所内

(74) 代理人 100064285

弁理士 佐藤 一雄 (外3名)

Fターム (参考) 5B075 ND03 PP24 PQ02 PQ46 PQ74

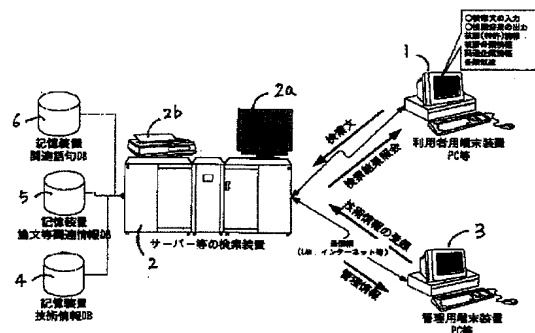
PRO6

(54) 【発明の名称】 技術情報検索システム

(57) 【要約】

【課題】 熟練を要することなく、簡易な手順で高精度の検索が可能な技術情報検索システムを提供する。

【解決手段】 本発明の技術情報検索システムは、検索文の入力や検索結果の出力を行う利用者用端末装置1と、検索文に基づいて技術情報の検索処理を行う検索装置2と、検索装置2に技術情報の登録を行う管理用端末装置3とを備えている。ユーザが入力した検索文を形態素解析して単語ベクトルを生成し、技術情報データベース装置4に登録されている各技術情報の単語ベクトルとの内積を計算した結果に基づいて類似度を判定するため、各技術情報との類似度を簡易かつ精度よく判定することができる。また、類似度の高低がわかるような形態で検索結果を表示するため、複数の技術情報が検索されても、各技術情報の類似度の高さを客観的に把握することができる。



【特許請求の範囲】

【請求項 1】 入力された検索文の中から関連語を抽出する関連語抽出手段と、

抽出された関連語に対して形態素解析を行う第 1 の形態素解析手段と、

前記第 1 の形態素解析手段による形態素解析により得られた単語それぞれに対して、技術情報の内容を加味した重み付けを行ってベクトル情報を生成する第 1 のベクトル生成手段と、

複数の検索対象技術情報に対して形態素解析を行う第 2

の形態素解析手段と、  
前記第 2 の形態素解析手段による形態素解析により得られた単語それぞれに対して、技術情報の内容を加味した重み付けを行ってベクトル情報を生成する第 2 のベクトル生成手段と、

前記第 1 のベクトル生成手段で生成されたベクトル情報と前記第 2 のベクトル生成手段で生成されたベクトル情報との間で内積を演算する内積演算手段と、

前記内積演算手段の演算結果に基づいて、検索文と検索対象技術情報との類似度を判定する類似度判定手段と、  
前記類似度判定手段による判定結果の出力対象を選択する出力対象選択手段と、

選択された出力対象を前記類似度判定手段により判定された類似度順に並べて出力する検索結果出力手段とを備えることを特徴とする技術情報検索システム。

【請求項 2】 前記出力対象選択手段により選択される出力対象は、技術情報別、関連技術分類別および関連企業別の少なくとも一つを含むことを特徴とする請求項 1 に記載の技術情報検索システム。

【請求項 3】 前記検索結果出力手段が出力する検索結果に特定の技術情報が含まれている回数と、その特定の技術情報に対する前記類似度判定手段による類似度判定結果と、その特定の技術情報に対する前記詳細情報表示制御手段による表示回数とに基づいて、各技術情報の市場価値を推定する市場価値推定手段をさらに備えることを特徴とする請求項 1 または 2 に記載の技術情報検索システム。

【請求項 4】 前記検索結果出力手段は、選択された出力対象を前記類似度判定手段により判定された類似度順に並べて出力する際、この出力対象の類似度が視覚的に把握可能な指標を表示させることを特徴とする請求項 1 ～ 3 のいずれかに記載の技術情報検索システム。

【請求項 5】 入力された検索文の中から関連語を抽出する関連語抽出手段と、

抽出された関連語に対して形態素解析を行う第 1 の形態素解析手段と、

前記第 1 の形態素解析手段による形態素解析により得られた単語それぞれに対して、技術情報の内容を加味した重み付けを行ってベクトル情報を生成する第 1 のベクトル生成手段と、

複数の検索対象技術情報に対して形態素解析を行う第 2 の形態素解析手段と、

前記第 2 の形態素解析手段による形態素解析により得られた単語それぞれに対して、技術情報の内容を加味した重み付けを行ってベクトル情報を生成する第 2 のベクトル生成手段と、

前記第 2 のベクトル生成手段で生成されたベクトル情報を格納する第 1 の情報蓄積手段と、

広告情報に対して形態素解析を行う第 3 の形態素解析手段と、

前記第 3 の形態素解析手段による形態素解析により得られた単語それぞれに対して、技術情報の内容を加味した重み付けを行ってベクトル情報を生成する第 3 のベクトル生成手段と、

前記第 3 のベクトル生成手段で生成されたベクトル情報を格納する第 2 の情報蓄積手段と、

前記第 1 の情報蓄積手段に格納されたベクトル情報と前記第 1 のベクトル生成手段で生成されたベクトル情報との間で内積を演算し、かつ前記第 2 の情報蓄積手段に格納されたベクトル情報と前記第 1 のベクトル生成手段で生成されたベクトル情報との間で内積を演算する内積演算手段と、

前記内積演算手段の演算結果に基づいて類似度を判定する類似度判定手段と、

判定された類似度に基づいて、類似度順に検索対象技術情報を出力するとともに、類似度が所定量よりも高い広告情報を出力する検索結果出力手段と、を備えることを特徴とする技術情報検索システム。

【請求項 6】 前記検索結果出力手段は、検索文との類似度が高い広告情報ほど大きく表示させることを特徴とする請求項 5 に記載の技術情報検索システム。

【請求項 7】 表示された広告情報の少なくとも一つを選択する広告選択手段と、

前記広告選択手段により選択された広告情報の詳細情報を表示させる詳細広告表示制御手段を備えることを特徴とする請求項 5 または 6 に記載の技術情報検索システム。

【請求項 8】 表示画面に特定の広告が表示された回数と、表示画面に表示された前記特定の広告と検索文との類似度と、前記詳細広告表示制御手段により前記特定の広告が選択された回数とに基づいて、前記特定の広告の掲載料金額を設定する掲載料金設定手段を備えることを特徴とする請求項 5 ～ 7 のいずれかに記載の技術情報検索システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、入力された検索文に基づいて技術情報の検索を行う技術情報検索システムに関する。

【0002】

【従来の技術】最近、特許情報を初めとして、大量の技術情報が電子化されてインターネット等を介してアクセス可能になってきた。技術情報の中には、無料で提供されるものや、特定のIDやパスワードで照合を行って有料で提供されるものがあり、また技術情報の言語やデータ形式もさまざまである。このため、これら大量の技術情報の中から所望の技術情報を簡易かつ迅速に検索する技術が必要になる。

【0003】

【発明が解決しようとする課題】特許情報などの技術情報の検索を行う場合、従来はキーワードを指定して検索を行うことが多かった。ところが、的確なキーワードを入力しないと検索の精度が上がらないという問題がある。

【0004】例えば、汎用的なキーワードを指定して検索を行うと、該当件数が急激に増加して、真に必要な情報の抽出が困難になってしまう。したがって、検索の精度を上げるには、熟練したサーチャーの存在が不可欠であり、一般の人が検索を行っても、精度の悪い検索しかできないことから、検索システム自体があまり利用されないという悪循環に陥っていた。

【0005】一方、インターネットの普及に伴い、ホームページ上に広告情報を表示する例が増えてきた。その例の一つとして、ユーザが入力した検索条件に対応する検索結果を表示する際に、その検索条件に関連のある広告情報を表示する広告付与システムが提案されている。

【0006】このシステムは、予め広告主が広告対象である製品やサービスの分類をシステム管理上の分類の中から選択して登録しておき、検索結果を表示する際に、その検索結果の分類に属する広告情報を同時に表示するものである。

【0007】しかしながら、広告対象である製品やサービスと、分類との適合度を上げるためには、分類の階層構造をより細かくしなければならぬ。このため、広告主側の負担が急激に増加し、広告件数そのものが減少してしまう。

【0008】また、同一分類に属する製品やサービスの情報が多数存在する場合、表示の優先度を決定する合理的な手段が提案されていない。

【0009】さらに、広告主は、自己の製品やサービスに関する広告情報がなるべく多く表示されるよう、広範囲の分類を指定したがることも多いと思われるが、そのようにすると、ユーザに大量の広告情報を提供することになり、広告情報を分類分けするメリットが薄れてしまう。

【0010】また、従来、特許のような技術情報の市場価値を評価する手法としては、当事者間で締結した契約に基づくライセンス料により評価するのが主流であった。しかしながら、ライセンス契約の内容は当事者以外には非公開である場合が多く、限られた分野の既存取引

についての評価しかできず、新しい技術の市場価値の予測手法としては適用できなかった。

【0011】本発明は、このような点に鑑みてなされたものであり、その目的は、熟練を要することなく、簡易な手順で高精度の検索が可能な技術情報検索システムを提供することにある。

【0012】

【課題を解決するための手段】上述した課題を解決するために、本発明は、入力された検索文の中から関連語を抽出する関連語抽出手段と、抽出された関連語に対して形態素解析を行う第1の形態素解析手段と、前記第1の形態素解析手段による形態素解析により得られた単語それぞれに対して、技術情報の内容を加味した重み付けを行ってベクトル情報を生成する第1のベクトル生成手段と、複数の検索対象技術情報に対して形態素解析を行う第2の形態素解析手段と、前記第2の形態素解析手段による形態素解析により得られた単語それぞれに対して、技術情報の内容を加味した重み付けを行ってベクトル情報を生成する第2のベクトル生成手段と、前記第1のベクトル生成手段で生成されたベクトル情報と前記第2のベクトル生成手段で生成されたベクトル情報との間で内積を演算する内積演算手段と、前記内積演算手段の演算結果に基づいて、検索文と検索対象技術情報との類似度を判定する類似度判定手段と、前記類似度判定手段による判定結果の出力対象を選択する出力対象選択手段と、選択された出力対象を前記類似度判定手段により判定された類似度順に並べて出力する検索結果出力手段とを備える。

【0013】本発明では、検索文の中から抽出した関連語に対応するベクトル情報と検索対象技術情報のベクトル情報との類似度を判定し、類似度順に並べて検索結果を出力するようにしたため、検索文とどの程度類似しているかを定量的に把握することができ、検索文に類似する技術情報を簡易かつ迅速に検索することができる。

【0014】また、本発明は、入力された検索文の中から関連語を抽出する関連語抽出手段と、抽出された関連語に対して形態素解析を行う第1の形態素解析手段と、前記第1の形態素解析手段による形態素解析により得られた単語それぞれに対して、技術情報の内容を加味した重み付けを行ってベクトル情報を生成する第1のベクトル生成手段と、複数の検索対象技術情報に対して形態素解析を行う第2の形態素解析手段と、前記第2の形態素解析手段による形態素解析により得られた単語それぞれに対して、技術情報の内容を加味した重み付けを行ってベクトル情報を生成する第2のベクトル生成手段と、前記第2のベクトル生成手段で生成されたベクトル情報を格納する第1の情報蓄積手段と、広告情報に対して形態素解析を行う第3の形態素解析手段と、前記第3の形態素解析手段による形態素解析により得られた単語それぞれに対して、技術情報の内容を加味した重み付けを行っ

てベクトル情報を生成する第3のベクトル生成手段と、前記第3のベクトル生成手段で生成されたベクトル情報を格納する第2の情報蓄積手段と、前記第1の情報蓄積手段に格納されたベクトル情報と前記第1のベクトル生成手段で生成されたベクトル情報との間で内積を演算し、かつ前記第2の情報蓄積手段に格納されたベクトル情報と前記第1のベクトル生成手段で生成されたベクトル情報との間で内積を演算する内積演算手段と、前記内積演算手段の演算結果に基づいて類似度を判定する類似度判定手段と、判定された類似度に基づいて、類似度順に検索対象技術情報を出力するとともに、類似度が所定量よりも高い広告情報を出力する検索結果出力手段と、を備える。

【0015】また、本発明では、検索文と検索対象技術情報との類似度を考慮に入れて広告情報を提供するようにしたため、関心のある者のみに広告情報を提供でき、広告効果を高めることができる。

【0016】

【発明の実施の形態】以下、本発明に係る技術情報検索システムについて、図面を参照しながら具体的に説明する。

【0017】（第1の実施形態）図1は本発明に係る技術情報検索システムの第1の実施形態の概略構成を示すブロック図である。図1の技術情報検索システムは、検索文の入力や検索結果の出力を行う利用者用端末装置1と、検索文に基づいて技術情報の検索処理を行う検索装置2と、検索装置2に技術情報の登録を行う管理用端末装置3とを備えている。

【0018】各端末装置1はパーソナルコンピュータ（以下、PC）などの比較的小型のコンピュータで構成され、検索装置2はサーバなどの比較的大型のコンピュータで構成される。

【0019】検索装置2は、表示装置2aとプリンタ2bとを有し、検索装置2には、検索対象となる技術情報が登録された技術情報データベース（DB）装置4と、論文等の関連情報が登録された関連情報データベース装置5と、関連語句が登録された関連語句データベース装置6とが接続されている。

【0020】各端末装置1と検索装置2とは、ローカルエリアネットワーク（以下、LAN）、インターネットまたは通信回線等の通信網を介してデータの送受信を行う。

【0021】ユーザは、利用者用端末装置1のキーボード等から検索文を入力する。検索文には、字数制限や文字種制限がなく、各自の言葉で任意の文章を入力することができる。また、検索対象の正式な名称が不明でも、概念や仕組みを説明する文章を入力することにより検索を行うことができる。例えば、「水素と酸素を用いて、発電する仕組みや技術」という文章を検索文として入力することで検索を行える。この場合、後述するように、

「燃料電池」に関する技術情報が抽出される。

【0022】検索装置2は、利用者用端末装置1で入力された検索文に基づいて検索処理を行い、その検索結果は利用者用端末装置1に送られる。また、管理用端末装置3は検索装置2に技術情報を登録し、検索装置2は管理用端末装置3に管理情報を送信する。

【0023】図2は図1の技術情報検索システムの第1の実施形態の処理動作を説明するフローチャートである。ユーザが利用者用端末装置1から検索文を入力する（ステップS1）と、その検索文は検索装置2に送られる。検索装置2は、検索文に含まれる関連語を抽出する（ステップS2）。

【0024】続いて、検索装置2は、形態素解析を行う（ステップS3）。形態素解析とは、文章（自然文）に含まれている単語を抽出するものであり、具体的には複合名詞句と数詞を抽出する。例えば、上述した「水素と酸素を用いて、発電する仕組みや技術」という検索文の場合、「水素」、「酸素」、「発電」、「仕組み」、「技術」、「用いる」の6つの単語とその類似語が抽出される。

【0025】続いて、形態素解析により抽出された単語それぞれ（検索語群）を、技術情報の内容を加味した重み（ウエイト）をつけて数値情報（ベクトル）化する（ステップS4）。この処理は検索語群のインデキシングと呼ばれており、この処理で生成される数値情報を以下では単語ベクトルと呼ぶ。例えば、データベース装置内のどのファイルにも頻出する用語や、接頭語・接尾語などは重要度が低いと判断してウエイトを低くし、特定の技術情報ファイルにのみ頻出する用語は、その技術情報を代表する単語とみなしてウエイトを高くする。

【0026】ステップS4の処理が終了すると、次に、技術情報データベース装置4に登録されている技術情報と検索語群のインデックスとの間でベクトル処理を行って類似度を検出する（ステップS5）。

【0027】ここでは、技術情報データベース装置4に登録されている技術情報をベクトル表現した単語ベクトルと検索文の検索語群をベクトル表現した単語ベクトルとを比較し、ベクトルの向きが近い情報を抽出する。ベクトルの向きの近さは内積で表すことができ、この内積の値を本実施形態では類似度という指標で表す。

【0028】例えば、上述した検索文に含まれる「水素」、「酸素」、「発電」、「仕組み」、「技術」、「用いる」の6つの単語の意味的な重み（ウエイト）がすべて等価の場合には、これらの単語の組み合わせに最も近い単語の組み合わせが含まれる文書を技術情報データベース装置4の中から検索する。

【0029】単語一つ一つが独立した変数だとすると、上記の検索文の場合、数ある次元のうち、特定の6次元の合成ベクトルで表される。このため、ステップS5では、この合成ベクトルの向きに最も近い、すなわち内積

が 1 に近い技術情報を検索する。

【0030】例えば、「燃料電池」の説明として、技術情報データベース装置 4 の中に、「物質が反応するとき生じる化学エネルギーを電気エネルギーとして取り出す装置。通常、水素と酸素の化学反応を用いるが、水素はメタノール等を改質することにより得る。」という技術情報があったとすると、この技術情報中の「電気」や「エネルギー」という単語から「発電」という単語が関連づけられる。また、検索文中に「水素」と「酸素」という単語があることから、上記の「燃料電池」に関する技術情報が高い類似度で抽出される。

【0031】ところで、図 2 のフローチャートにおいて、ステップ S 2 が関連語抽出手段に、ステップ S 3 が第 1 の形態素解析手段に、図 10 のステップ S 54 が第 2 の形態素解析手段に、図 2 のステップ S 5 が内積演算手段に、ステップ S 6 が類似度判定手段に、ステップ S 9 が出力対象選択手段に、ステップ S 10～S 12 が検索結果出力手段に、それぞれ対応する。

【0032】図 3 は上述した検索文に対する技術情報データベース装置 4 の検索手順を示す図である。技術情報データベース装置 4 に登録されるべき各技術情報は、形態素解析されて、各単語ごとに単語ベクトルが生成された後、技術情報データベース装置 4 に登録される（ステップ S 21～S 24）。

【0033】一方、ユーザが入力した検索文も、形態素解析されて、各単語ごとに単語ベクトルが生成される（ステップ S 25～S 27）。

【0034】続いて、検索文の単語ベクトルと技術情報データベース装置 4 に登録された各単語ベクトルとの間で内積を演算し（ステップ S 28）、その演算結果に基づいて類似度を判定する（ステップ S 29）。

【0035】なお、技術情報データベース装置 4 には、多数の技術情報が登録されており、各技術情報ごとに検索語群との間で内積を演算して類似度を検出する。

【0036】このように、本実施形態では、内積の大小により類似度を判定できるため、類似度の判定を簡易かつ精度よく行うことができる。

【0037】図 2 のステップ S 5 の処理が終了すると、続いて、ステップ S 5 で数学的に算出された類似度が予め定めた基準類似度より小さいか否かを判定する（ステップ S 6）。算出された類似度の中に基準類似度を超えるものがなければ、検索文に関連のある技術情報はないと判断して、その旨を利用者用端末装置 1 に表示する（ステップ S 7）。

【0038】一方、類似度が基準類似度以上と判定された技術情報については、その技術情報を技術情報データベース装置 4 から抽出する（ステップ S 8）。続いて、検索結果をどのような形式で利用者用端末装置 1 の画面に表示するかを選択する（ステップ S 9）。例えば、ユーザが技術情報別に表示する旨の選択を行った場合は、

類似度の高い技術情報から順に利用者用端末装置 1 に表示する処理を行う（ステップ S 10）。

【0039】図 4 は技術情報別に検索結果を表示する例を示す図である。画面内には、ユーザが入力した検索文を表示する領域 W 1 と、技術情報抽出結果リストを表示する領域 W 2 と、技術情報の詳細内容を表示する領域 W 3 とが設けられる。

【0040】技術情報抽出結果リストには、類似度の高い順に技術情報の名称がリスト表示される。また、各技術情報が検索文にどの程度類似しているかを一目で把握できるように、類似度が数値で表示される。類似度が高いほど、検索文に近い内容であることを示しており、検索者にとって有用な情報であることを示す指標である。

【0041】なお、類似度を数値で表示する代わりに、図 5 に示すように棒グラフ等で表示してもよい。グラフで表示すれば、各技術情報の類似度の差異を容易に把握することができる。

【0042】技術情報抽出結果リスト内の任意の技術情報をユーザがマウス等で選択すると、その技術情報に対応する詳細内容が右側の表示領域 W 3 に表示される。この領域には、ユーザが選択した技術情報の書誌の情報だけでなく、全文情報や関連情報も表示してもよい。

【0043】一方、図 2 のステップ S 6 において、ユーザが関連技術分類（例えば、国際特許分類 IPC）別に表示する旨の選択を行った場合は、類似度の高い関連技術分類から順に利用者用端末装置 1 に表示する処理を行う（ステップ S 8）。

【0044】図 6 は図 2 のステップ S 8 の処理の詳細フローチャートである。まず、ステップ S 7 と同様に、技術情報別の類似度検出を行う（ステップ S 31）。続いて、類似度検出された各技術情報の関連技術分類を検出する（ステップ S 32）。

【0045】続いて、関連技術分類別に類似度の平均値やばらつきを算定する（ステップ S 33）。続いて、類似度の高い順に関連技術分類を並べたリストを作成する（ステップ S 34）。

【0046】図 7 は図 2 のステップ S 8 に対応する検索結果表示画面の一例を示す図である。図 4 との違いは、関連技術分類別に検索結果を表示する点である。関連技術分類別に検索結果を集計して表示し、ユーザは、検索文に関連のある技術分類を把握でき、技術分類を指定してさらなる検索を行うことができる。

【0047】一方、図 2 のステップ S 6 において、ユーザが関連企業別に表示する旨の選択を行った場合は、平均類似度の高い関連企業から順に利用者用端末装置 1 に表示する処理を行う（ステップ S 9）。

【0048】図 8 はステップ S 9 の処理の詳細フローチャートである。技術情報別の類似度検出を行った（ステップ S 41）後、類似度検出された各技術情報の関連企業を検出する（ステップ S 42）。

【0049】続いて、関連企業別に類似度の平均値やばらつきを算定する(ステップS43)。続いて、関連企業を類似度の高い順に並べたリストを作成する(ステップS44)。

【0050】図9は図2のステップS9に対応する検索結果表示画面の一例を示す図である。図4との違いは、関連企業別に検索結果を集計して表示する点である。関連企業別に検索結果を表示することにより、ユーザは、検索文に関連のある企業名を一目で把握できる。

【0051】技術情報データベース装置4に登録されている技術情報が多いほど、利用価値が高まるため、本実施形態では、新たな技術情報を逐次、技術情報データベース装置4に登録できるようにしている。

【0052】図10は管理用端末装置3が行う技術情報登録処理の処理手順を示すフローチャートである。まず、登録すべき技術情報を取り込む(ステップS51)。続いて、取り込んだ技術情報のデータ形式を、技術情報データベース装置4に見合ったデータ形式に変換する(ステップS52)。ここでは、主に、タグ情報の変換を行う。

【0053】図11は技術情報データベース装置4に登録されている各技術情報をもつタグの一覧を示す図であり、特許文書用のタグの例を示している。

【0054】特許文書用のタグには、文書の先頭を示す<doc>、文書のタイトルを示す<title>、出願年を示す<createddate>、特許権者を示す<au>、キーワードを示す<keywordt>、本文を示す<body>、および文書の最後を示す</doc>などがある。

【0055】図12は技術情報データベース装置4に登録されている技術情報の一例を示す図であり、特許文書の例を示している。図示のように、各技術情報は、タグを付けた状態で技術情報データベース装置4に登録されるため、これらタグに基づいて検索を行うことにより、所望の技術文書を簡易かつ迅速に探り出すことができる。

【0056】図10のステップS52のタグ変換処理が終了すると、次に検索用の技術情報データを作成する(ステップS53)。続いて、図2のステップS3と同様の手法で形態素解析を行って関連語を抽出した(ステップS54)後、各関連語に重要度等を加味した重み付けを行う(ステップS55)。このとき、各関連語をベクトル的に表現した数値データを生成し、このデータを技術情報データベース装置4に登録する(ステップS56)。

【0057】図13～図15は図10のステップS52で行われるタグ変換処理の処理手順を示すフローチャートである。まず、エラー情報を記録するエラーログファイルをオープンする(ステップS71)。次に、処理開始時刻等の処理情報を記録する処理ログファイルをオープンする(ステップS72)。

【0058】次に、処理開始時刻を処理ログに記録する

(ステップS73)。次に、特許文書が記録されたCD-ROM内の特許文書データを順次読み出す(ステップS74)。次に、種々のタグ変数を初期化する(ステップS75)。具体的には、作者名を表す変数\$authorにはヌルデータを、キーワードを表す変数\$ktには<keywordt>を、タイトルを表す変数\$titleには<title>を、作成日を表す変数\$c\_dateには<createddate>を入力する。

【0059】次に、フラグを初期化した(ステップS76)後、特許データファイルをオープンし(ステップS77)、特許文書データを一行ずつ読み出す(ステップS78)。読み出した行に、隅付き括弧で囲まれた「発明の名称」という用語が存在するか否かを判定する(ステップS79)。この用語が存在すれば、変数\$titleに、発明の名称の具体名を代入する(ステップS80)。

【0060】ステップS79でNOと判定された場合、あるいはステップS80の処理が終了した場合には、読み出した行に、隅付き括弧で囲まれた「出願日」という用語が存在するか否かを判定する(ステップS81)。この用語が存在すれば、日付を表す全角文字をyy/mm/ddの形式の半角文字に変換した(ステップS82)後、この半角文字を変数\$c\_dateに代入する(ステップS83)。

【0061】ステップS81でNOと判定された場合、あるいはステップS83の処理が終了した場合には、読み出した行に、隅付き括弧で囲まれた「特許権者」または「出願人」という用語が存在するか否かを判定する(図14のステップS84)。この用語が存在すれば、隅付き括弧で囲まれた「氏名または名称」という用語が存在する行までスキップした後、その行に記載された具体的な氏名または名称を変数\$authorに代入する(ステップS85)。

【0062】ステップS84でNOと判定された場合、あるいはステップS85の処理が終了した場合には、隅付き括弧で囲まれた「国際特許分類」という用語が存在するか否かを判定し(ステップS86)、この用語が存在すれば、キーワードフィールド・フラグをオンする(ステップS87)。

【0063】ステップS86でNOと判定された場合、あるいはステップS87が終了した場合、特許データファイルの最終行までタグ変換を行ったか否かを判定する(ステップS88)。まだ最終行までタグ変換を行っていない場合には、ステップS78以降の処理を繰り返す。

【0064】最終行までタグ変換を行った場合には、タグ抜けが存在するか否かを判定する(ステップS89)。タグ抜けが存在する場合には、エラーログの出力を行う(ステップS90)。

【0065】ステップS89でNOと判定された場合、あるいはステップS90の処理が終了した場合には、出力ディレクトリが存在するか否かを判定する(ステップS91)。出力ディレクトリが存在しない場合にはディレクトリを初期作成する(ステップS92)。

【0066】ステップS91でYESと判定された場合、あるいはステップS92の処理が終了した場合、対象年ファイルをオープンする（ステップS93）。まず、開始タグ<doc>をオープンする（ステップS94）。次に、抽出したタグを対象年ファイルに出力する（ステップS95）。ここで、抽出するタグは、例えば、<title>、<createddate>、<au>、<keyword>である。

【0067】次に、特許文書本体のタグ<body>の後に、特許文書データの全内容を出力する（図15のステップS96）。次に、対象年ファイルに終了タグ</doc>を出力した後、同ファイルをクローズする（ステップS97）。

【0068】次に、CD-ROMに記録された全特許文書データについて、タグ変換と年別ファイルへの出力が終了したか否かを判定する（ステップS98）。次に、タグ変換と年別ファイルへの登録が終了したCD-ROMの名称や番号を処理ログに記録する（ステップS99）。

【0069】次に、すべてのCD-ROMについての処理が終了したか否かを判定する（ステップS100）。まだ処理が終了していないCD-ROMがあれば、ステップS74以降の処理を行う（ステップS101）。次に、処理ログファイルとエラーログファイルをクローズする（ステップS102）。

【0070】図16および図17は図2のフローチャートで説明した検索処理の詳細内容を示すフローチャートである。まず、検索文字列が記録されている入力ファイルをオープンした（ステップS121）後、検索結果を記録する出力ファイルをオープンする（ステップS122）。

【0071】次に、検索対象である技術情報を格納した参照文書数をセットした（ステップS123）後、検索文字列を読み込む（ステップS124）。次に、読み込んだ検索文字列にエラーがあるか否かを判定する（ステップS125）。エラーがない場合には、検索文字列のベクトルを生成する（ステップS126）。次に、生成したベクトルに基づいて技術情報の検索処理を行う（ステップS127）。

【0072】次に、ユーザが設定した関連度がゼロか否かを判定する（ステップS128）。関連度がゼロより大きければ、その関連度に基づいて関連語抽出を行う（ステップS129）。次に、抽出した関連語に基づいて、技術情報の検索を行う（ステップS130）。

【0073】ステップS128でYESと判定された場合、ステップS130の処理が終了した場合は、IPCコードの集計および整列を行った（ステップS131）後に、特許権者の集計および整列を行い（ステップS132）、その後、出願年の集計および整列を行い（ステップS133）、次に、ステップS130～S132の集計および整列結果を出力する（ステップS134）。

【0074】次に、ユーザが設定した要約率の数値がゼロか否かを判定する（ステップS135）。ここで、要約

率とは、全文章数に対する要約文章数の比率である。要約率の数値がゼロより大きい場合には、要約処理を行う（ステップS136）。

【0075】次に、すべての検索結果が得られたか否かを判定し（ステップS137）、まだ得られていないものがあれば、ステップS124以降の処理を繰り返す。すべての検索結果が得られた場合には、その結果を出力した（ステップS138）後、処理を終了する。

【0076】このように、第1の実施形態では、ユーザが入力した検索文を形態素解析して単語ベクトルを生成し、技術情報データベース装置4に登録されている各技術情報の単語ベクトルとの内積を計算した結果に基づいて類似度を判定するため、各技術情報との類似度を簡易かつ精度よく判定することができる。

【0077】また、類似度の高低がわかるような形態で検索結果を表示するため、複数の技術情報が検索されても、各技術情報の類似度の高さを客観的に把握することができる。

【0078】さらに、第1の実施形態では、検索文の中に技術的なキーワードが含まれていなくても、精度の高い検索を行えるため、通常の文章と同じような形式で検索文を作成でき、検索文の作成に熟練を要さなくなるため、ユーザの使い勝手が向上する。

【0079】（第2の実施形態）第2の実施形態は、検索結果を表示する際、検索文に関連のある広告情報を併せて表示するものである。

【0080】図18は本発明に係る技術情報検索システムの第2の実施形態の概略構成を示すブロック図である。図18では、図1と共通する構成部分には同一符号を付しており、以下では相違点を中心に説明する。

【0081】図18の技術情報検索システムは、製品（商品）やサービスの広告情報を提供する広告主用端末装置7を備えている。この広告主用端末装置7はPCなどで構成され、インターネット等の通信網を介して検索装置2とデータの送受信を行う。

【0082】また、図18の検索装置2は、広告主用端末装置7から提供された広告情報を登録する広告情報データベース装置（広告情報DB）9を有する。このデータベース装置9は、上述した技術情報データベース装置4や関連情報データベース装置5と同じ記録装置に格納されてもよいし、それぞれ別個に記録装置を設けてもよい。

【0083】図19は図18の技術情報検索システムの第2の実施形態の概略構成を示すフローチャートである。図19には、新たな技術情報を技術情報データベース装置4に登録する手順と、検索手順と、新たな広告情報を広告情報データベース装置9に登録する手順とが示されている。これらのうち、技術情報データベース装置4に登録する手順（ステップS151～S156）は図10と同じであるため、説明を省略する。

【0084】広告主が広告主端末装置を介して製品やサービスの広告情報を入力すると（ステップS157）、検索装置2はその入力情報を形態素解析し（ステップS158）、重要度等を加味して広告情報の重み付け処理（インデキシング）を行う（ステップS159）。より具体的には、広告情報に関する単語ベクトルを生成して、広告情報データベース装置9に登録する（ステップS160）。

【0085】なお、広告主は、新たな広告情報を登録する場合には、図15に示すフォーマットに従って必要な事項を記録した文書を、通信網を介して検索装置に送信する。

【0086】一方、ユーザが利用者用端末装置1を介して検索文を入力すると（ステップS161）、検索装置2はその検索文を形態素解析し（ステップS162、S163）、重要度等を加味して検索語群の重み付け処理を行う（ステップS164）。

【0087】続いて、検索装置2は、検索語群に関する単語ベクトルと技術情報データベース装置4に登録されている各単語ベクトルとの間で内積を演算するとともに、検索語に関する単語ベクトルと広告情報データベース装置9に登録されている各単語ベクトルとの間で内積を演算して類似度を検出する（ステップS165）。

【0088】続いて、検索装置2は、内積の演算結果に基づいて類似度を判定する（ステップS166）。より具体的には、内積の演算結果が予め入力された基準類似度よりも小さいか否かを判定する。内積の演算結果すべてが基準類似度より小さい場合には、類似度は低いと判断して、該当する技術情報がない旨の表示を行う（ステップS167）。

【0089】一方、内積の演算結果のうち、基準類似度以上のものがあれば、該当する技術情報を技術情報データベース装置4から抽出し、かつその技術情報に関連のある広告情報を広告情報データベース装置9から抽出する（ステップS168）。

【0090】続いて、検索結果の表示形式をユーザに選択させた（ステップS169）後、技術情報と、その技術情報に関連のある広告情報を、技術情報別、関連技術分類別、または関連企業別に表示する（ステップS170、S171、S172）。

【0091】ところで、図19のステップS162が関連語抽出手段に、ステップS163が第1の形態素解析手段に、ステップS154が第2の形態素解析手段に、ステップS158が第3の形態素解析手段に、ステップS156が第1の情報蓄積手段に、ステップS160が第2の情報蓄積手段に、ステップS165が内積演算手段に、ステップS166が類似度判定手段に、ステップS170～S172が検索結果出力手段に、それぞれ対応する。

【0092】図20は利用者用端末装置1に表示される検索結果の画面表示例を示す図である。図示のように、

検索文の表示領域、技術情報の抽出結果リストの表示領域、および技術情報の詳細内容の表示領域以外に、広告情報の表示領域が設けられる。この表示領域のサイズは、広告情報の数に応じて任意に変更してもよい。

【0093】広告情報を表示する際、類似度の高低に応じて広告の表示サイズを変えるようにすれば、ユーザの関心の高い広告情報ほど大きく表示することができ、広告の効果をよりいっそう高めることができる。また、ユーザにとっても、自分が関心のある製品またはサービスの広告情報が大きく表示されることから、この広告情報を確認することで、容易に検索目的を達成することができる。

【0094】図20の画面内に表示された特定の広告情報を、ユーザがマウス等で選択すると、その広告情報に関する製品またはサービスの概要を表示させたり、あるいは、その広告情報を提供する企業等のホームページが表示されるようにするのが望ましい。これにより、ユーザが必要とする場合のみ広告情報の詳細内容を表示できる。

【0095】ところで、広告掲載による定量的効果は、以下の（1）式で表される。

$$【0096】a = (b \times c) / \Sigma b \times d \quad \dots (1)$$

ここで、aは広告効果指標、bは画面に当該広告が表示された回数、cは当該広告が表示された際の類似度、dは当該広告のバナーがマウス等で選択されて当該広告の概要が参照された回数である。

【0097】（1）式にて算出されるaの値が大きいかほど広告効果が大きいことを示している。したがって、このaの値に基づいて広告掲載料金を設定すれば、各広告主に対して広告効果に応じた請求を行うことができる。上記の（1）式が掲載料金設定手段に相当する。

【0098】また、技術情報データベース装置4に登録されている技術情報の客観的な市場価値についても、同様に定量的に計測することができ、この場合の定量的効果は、以下の（2）式で表される。

$$【0099】S = (t \times u) / \Sigma t \times v \quad \dots (2)$$

ここで、Sは抽出技術情報の市場価値指標、tは画面に当該技術情報がリスト表示された回数、当該技術情報が表示された際の類似度、vはリストアップされた当該技術情報がマウス等により選択されて、当該技術情報の概要が参照された回数である。

【0100】このSは実際にライセンス取引が行われるときの価値基準として利用することができる。このSの値が大きいかほど価値が高いと考えられるため、Sの大きい技術情報については、タイトル情報を強調表示して、ユーザの注意を喚起するのが望ましい。上記の（2）式が市場価値推定手段に相当する。

【0101】また、広告主は、インターネット等を介して、オンライン上で広告掲載の申し込みをすることができる。図21はオンライン上での広告掲載の申し込み画

面の一例を示す図である。図示のように、広告主は、キャッチコピー、製品名、会社名、連絡先および製品概要などを各欄に入力した後、インターネット等を介して広告掲載の申し込みを行う。検索装置2は、広告主からの広告情報を広告情報データベース装置9に格納する。

【0102】このように、第2の実施形態では、ユーザが入力した検索文に関連のある技術情報を類似度順に表示するだけでなく、その技術情報に関連のある広告情報を表示するようにしたため、特定の製品やサービスに関心のあるユーザのみに広告を提供することができ、広告の効果を高めることができる。すなわち、検索結果に関連のある広告情報をユーザに提供するため、広告情報に関心があると思われるユーザのみに広告情報を提供でき、不特定多数の者に広告を行うよりも、広告による商品やサービスの売れ行きを伸ばせる。

【0103】また、検索結果が得られた段階でユーザに広告情報を提供するようにしたため、広告情報を無駄にユーザに提供しなくてすみ、ユーザに迷惑をかけなくなる。

【0104】さらに、オンラインにて広告情報の申し込みを行えるため、広告主の便宜が図れるとともに、広告情報の変更も容易に行える。

【0105】本実施形態では、広告情報の新規登録を行う場合は、広告主用端末装置7に図21に示す製品情報インデックスを表示させるようにしている。図21の製品情報インデックスは、キャッチコピー、製品またはサービス名、会社名、連絡先、および製品概要の各欄を有する。製品情報インデックスに入力された情報は、検索装置2に送られて広告情報データベース装置9に登録される。

【0106】

【発明の効果】以上詳細に説明したように、本発明によれば、検索文と検索対象技術情報との類似度を考慮に入れて検索結果を出力するようにしたため、個々の検索対象技術情報が検索文にどの程度類似しているかを容易かつ正確に把握することができる。これにより、先行技術文献をもとに特許審査を行う審査官の審査業務の迅速・効率化が図れる。また、企業の特許動向分析（例えば、研究開発戦略立案や他社の開発状況の調査など）の精度向上や効率化が図れるとともに、技術連関分析の実用化が図れる。

【0107】さらに、本発明によれば、検索文との類似度が高い広告情報を自動的に提供するようにしたため、広告に関心のある者のみに広告を提供でき、ダイレクトマーケティングによる対費用効果の高い広告掲載を実現できる。また、広告に関心のない者には広告情報が提供されなくなるため、検索サービスのユーザに迷惑をかけなくなる。

【0108】また、検索サービスに広告情報を含める分だけ、検索サービスの利用料を安価に設定でき、検索サ

ービスの利用促進が図れる。また、類似度に応じて広告情報の内容等を変えることにより、広告効果を高めることができる。

【図面の簡単な説明】

【図1】本発明に係る技術情報検索システムの第1の実施形態の概略構成を示すブロック図。

【図2】図1の技術情報検索システムの第1の実施形態の処理動作を説明するフローチャート。

【図3】上述した検索文に対する技術情報データベース装置の検索手順を示す図。

【図4】技術情報別に検索結果を表示する例を示す図。

【図5】技術情報抽出結果リストを示す図。

【図6】図2のステップS8の処理の詳細フローチャート。

【図7】図2のステップS8に対応する検索結果表示画面の一例を示す図。

【図8】ステップS9の処理の詳細フローチャート。

【図9】図2のステップS9に対応する検索結果表示画面の一例を示す図。

【図10】管理用端末装置3が行う技術情報登録処理の処理手順を示すフローチャート。

【図11】技術情報データベース装置に登録されている各技術情報をもつタグの一覧を示す図。

【図12】技術情報データベース装置に登録されている技術情報の一例を示す図。

【図13】図10のステップS52で行われるタグ変換処理の処理手順を示すフローチャート。

【図14】図13に続くフローチャート。

【図15】図14に続くフローチャート。

【図16】図2のフローチャートで説明した検索処理の詳細内容を示すフローチャート。

【図17】図16に続くフローチャート。

【図18】本発明に係る技術情報検索システムの第2の実施形態の概略構成を示すブロック図。

【図19】図18の技術情報検索システムの第2の実施形態の概略構成を示すフローチャート。

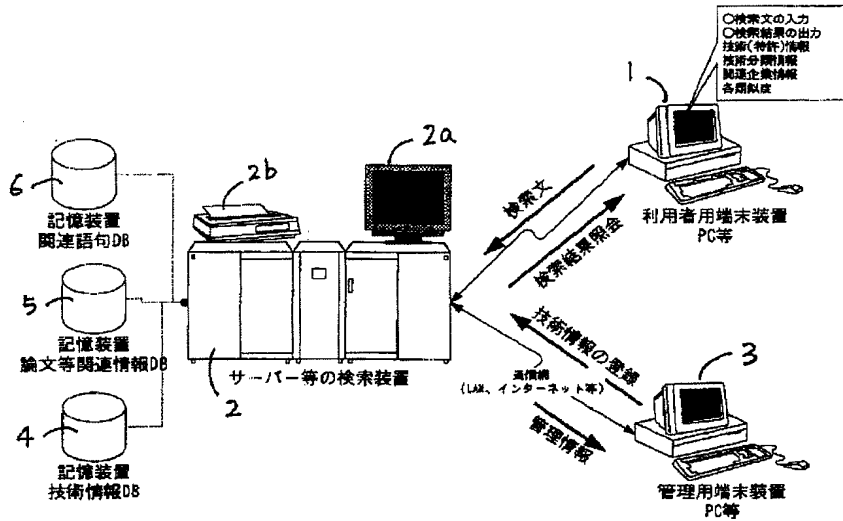
【図20】利用者用端末装置に表示される検索結果の画面表示例を示す図。

【図21】オンライン上での広告掲載の申し込み画面の一例を示す図。

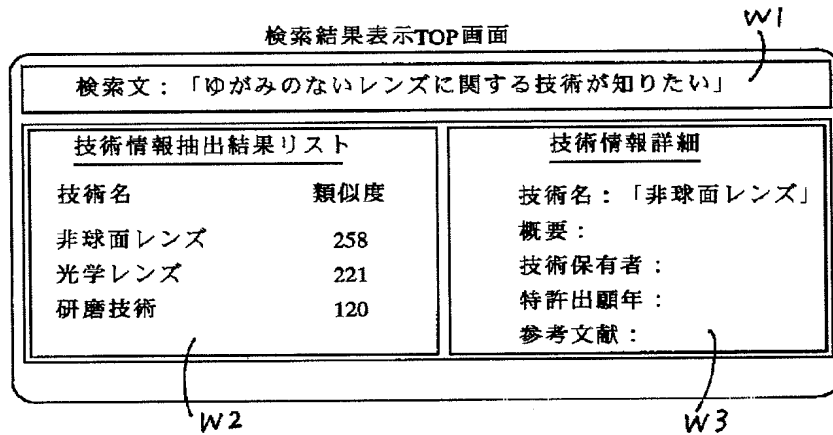
【符号の説明】

- 1 利用者用端末装置
- 2 検索装置
- 3 管理用端末装置
- 4 技術情報データベース装置
- 5 関連情報データベース装置
- 6 関連語句データベース装置
- 7 広告主用端末装置
- 8 製品情報データベース装置
- 9 広告情報データベース装置

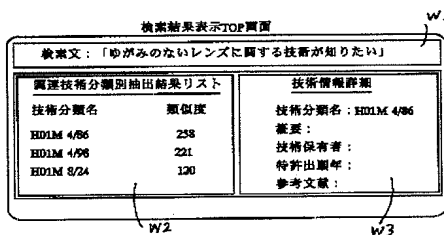
【図1】



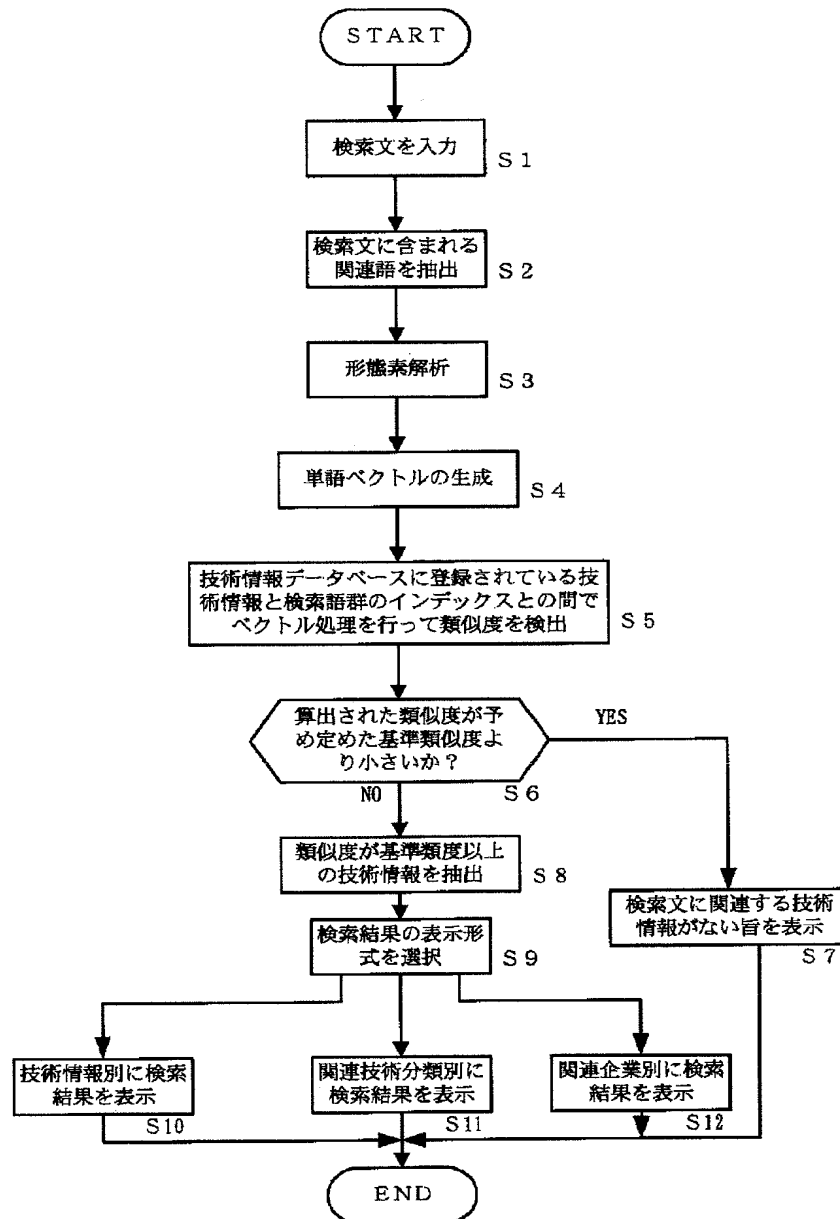
【図4】



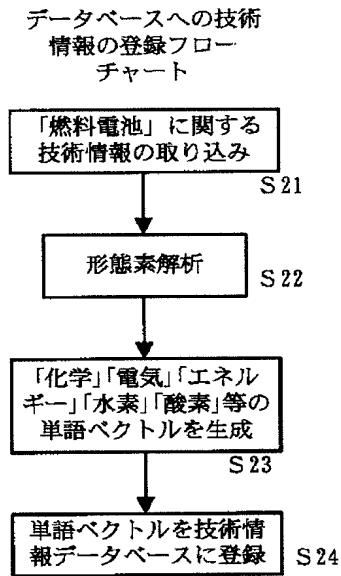
【図7】



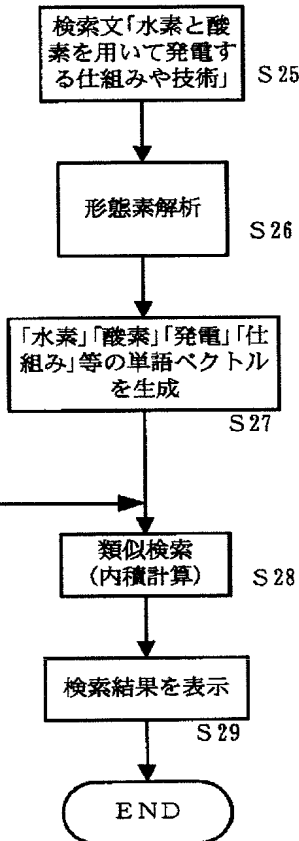
【図2】



【図3】



検索フローチャート



【図11】

本検索システムで付与するタグの種類

文書ははじめのタグ &lt;doc&gt;

タイトル &lt;title¥&gt;

出願年 &lt;createddate¥&gt;

特許権者 &lt;au¥&gt;

キーワード &lt;keywordt¥&gt;

内容(全内容) &lt;body&gt;

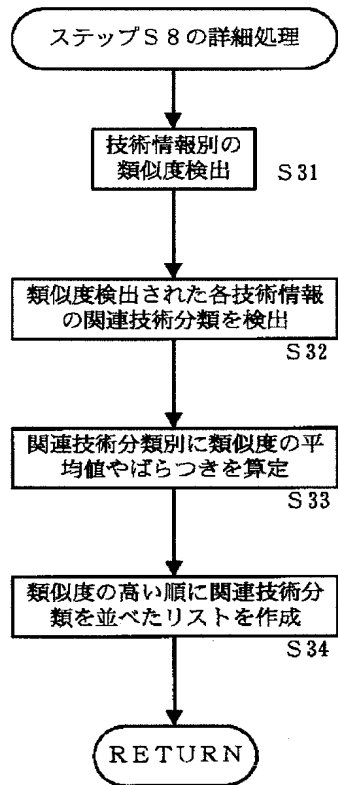
文書終わりのタグ &lt;/doc&gt;

【図5】

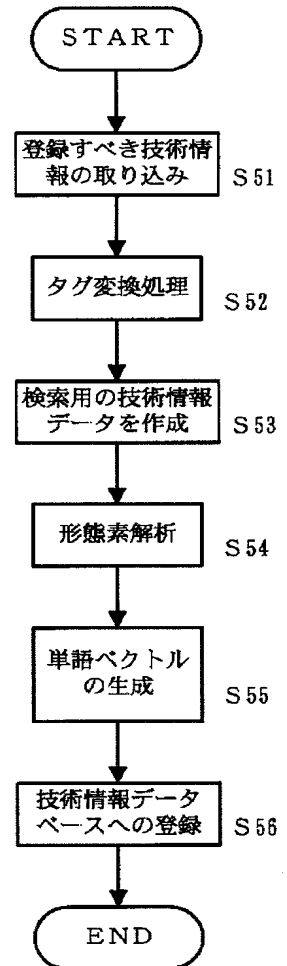
検索結果表示TOP画面

検索文：「ゆがみのないレンズに関する技術が知りたい」									
<p><u>技術情報抽出結果リスト</u></p> <table> <tr> <th>技術名</th> <th>類似度</th> </tr> <tr> <td>非球面レンズ</td> <td><div></div></td> </tr> <tr> <td>光学レンズ</td> <td><div></div></td> </tr> <tr> <td>研磨技術</td> <td><div></div></td> </tr> </table>	技術名	類似度	非球面レンズ	<div></div>	光学レンズ	<div></div>	研磨技術	<div></div>	<p><u>技術情報詳細</u></p> <p>技術名：「非球面レンズ」</p> <p>概要：</p> <p>技術保有者：</p> <p>特許出願年：</p> <p>参考文献：</p>
技術名	類似度								
非球面レンズ	<div></div>								
光学レンズ	<div></div>								
研磨技術	<div></div>								

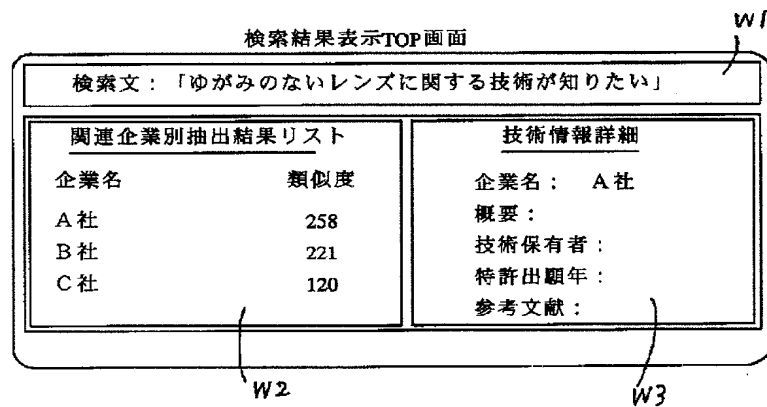
【図6】



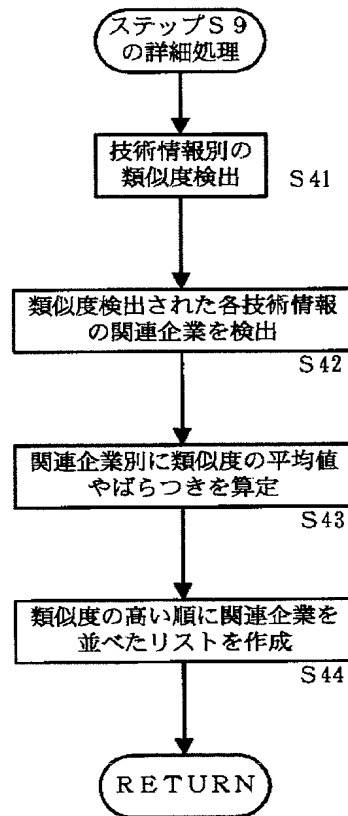
【図10】



【図9】



【図8】



【図12】

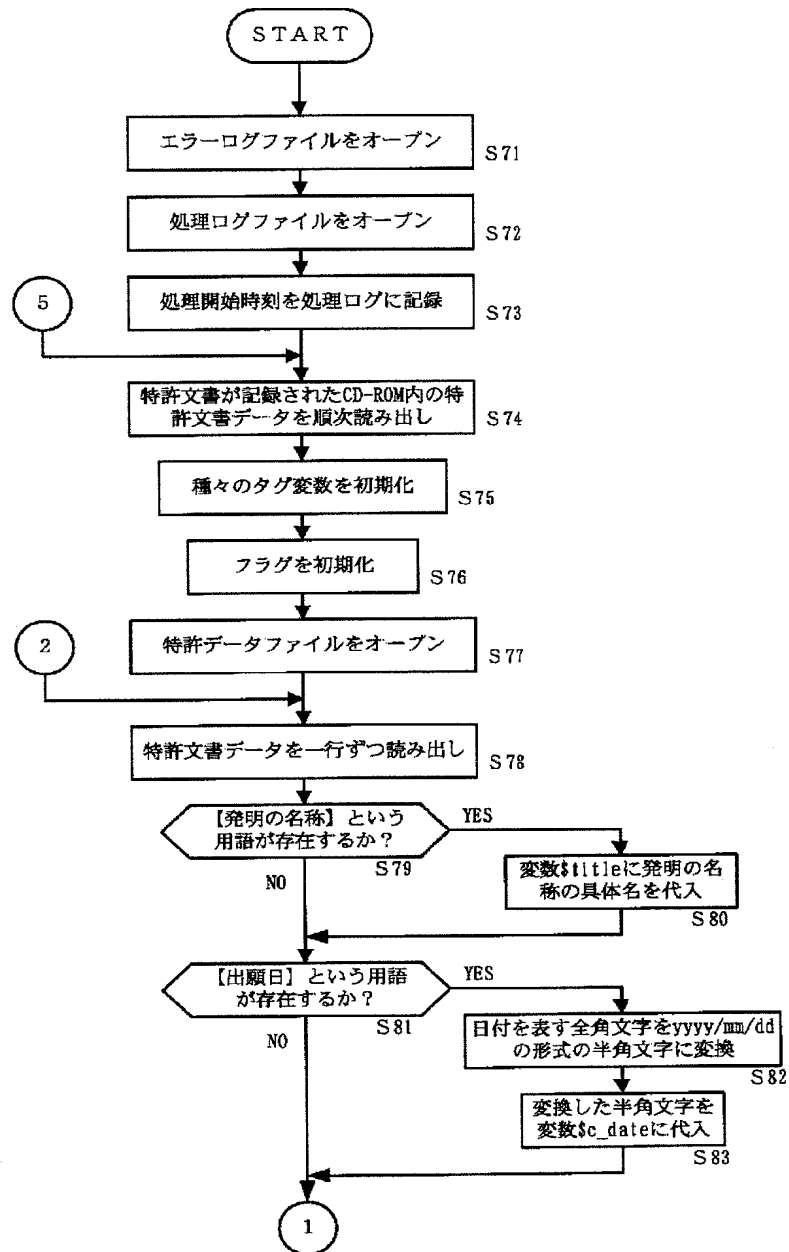
## 特許明細書のタグをつけた事例

```

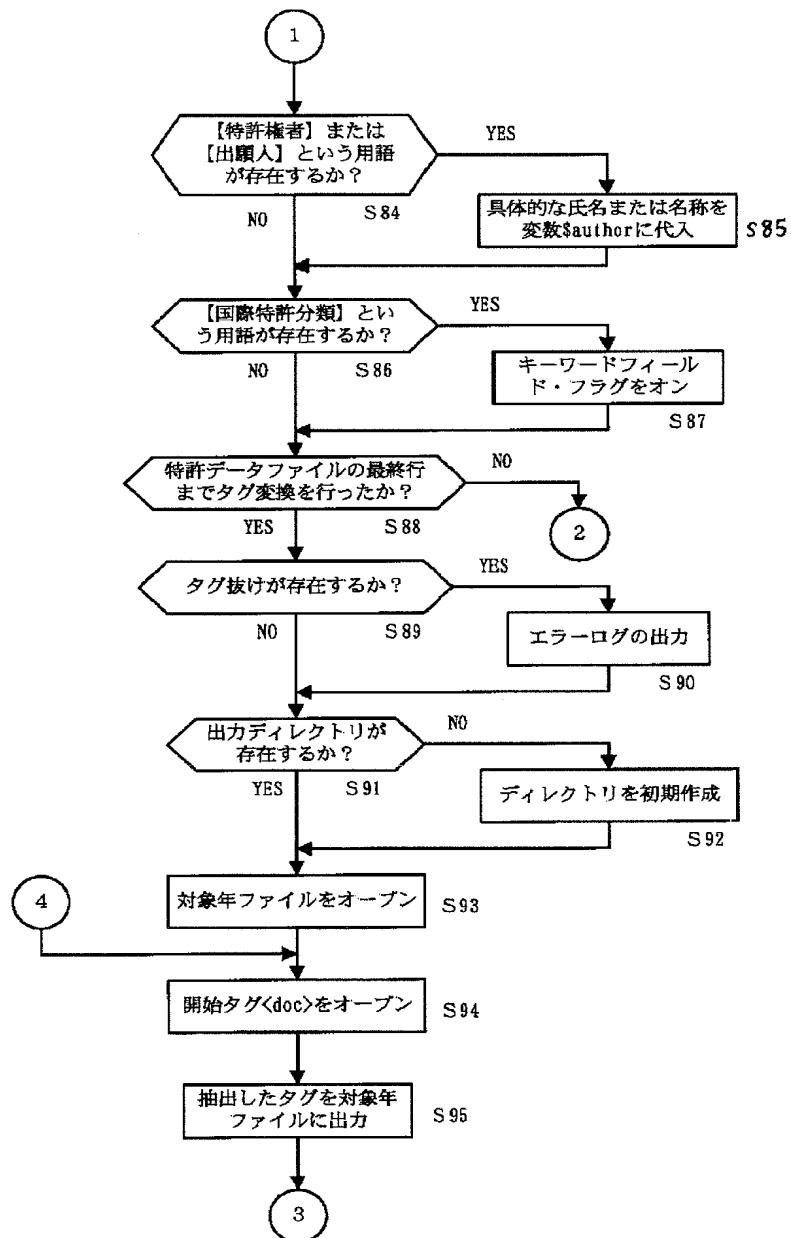
<doc>
<title>テーブリール
<createddate>1974/07/26
<au>ソニー株式会社
<keyword>, B65H 75/14      A 7030-3F, G11B 23/087  104  7177-5D
<body>
(19)【発行国】 日本国特許庁 (JP)
(12)【公報種別】 特許公報 (B2)
(11)【公告番号】 特公平6-6466
(24)(44)【公告日】 平成6年(1994)1月26日
(54)【発明の名称】 テーブリール
(51)【国際特許分類第5版】
      B65H 75/14      A 7030-3F
      G11B 23/087  104  7177-5D
【発明の数】 1
【全頁数】 4
～途中省略～
【第3図】
【第4図】
【第1図】
【第2図】
【第5図】
</doc>

```

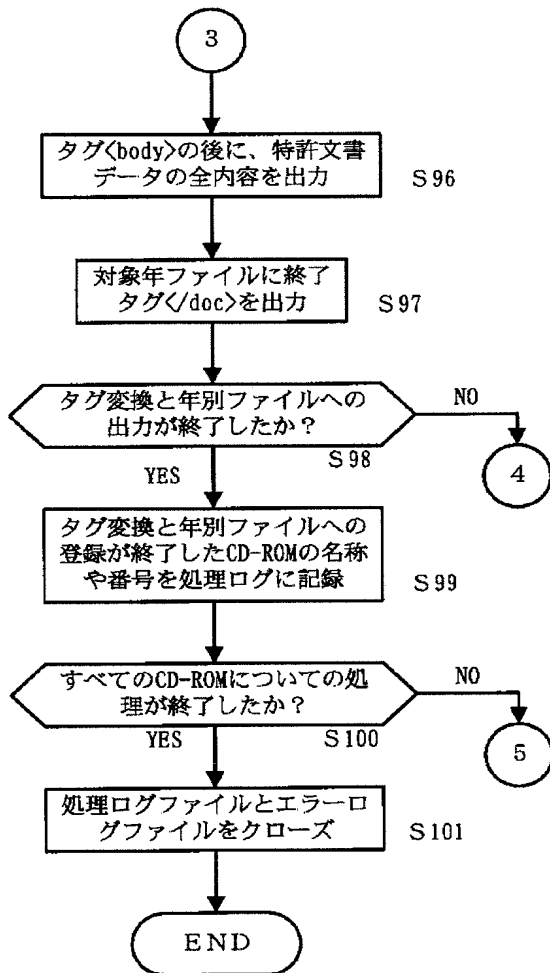
【図13】



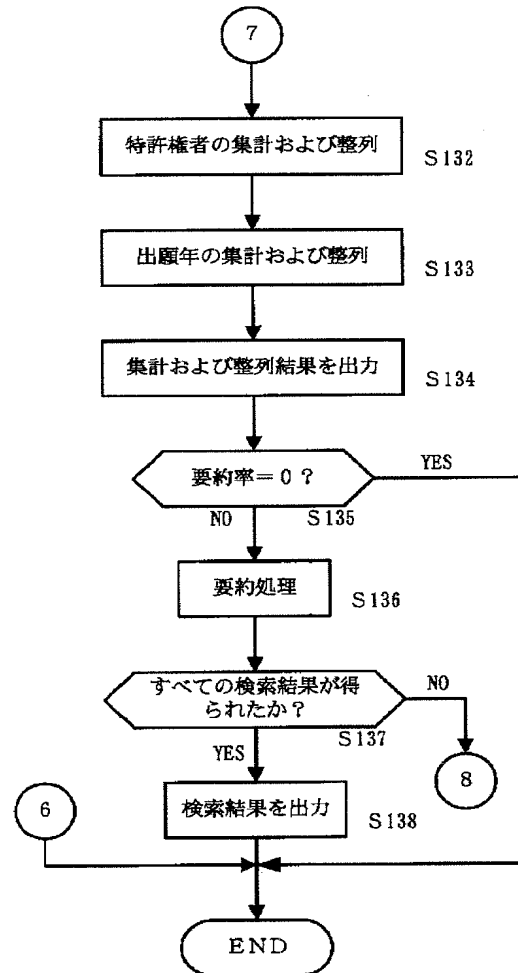
【図14】



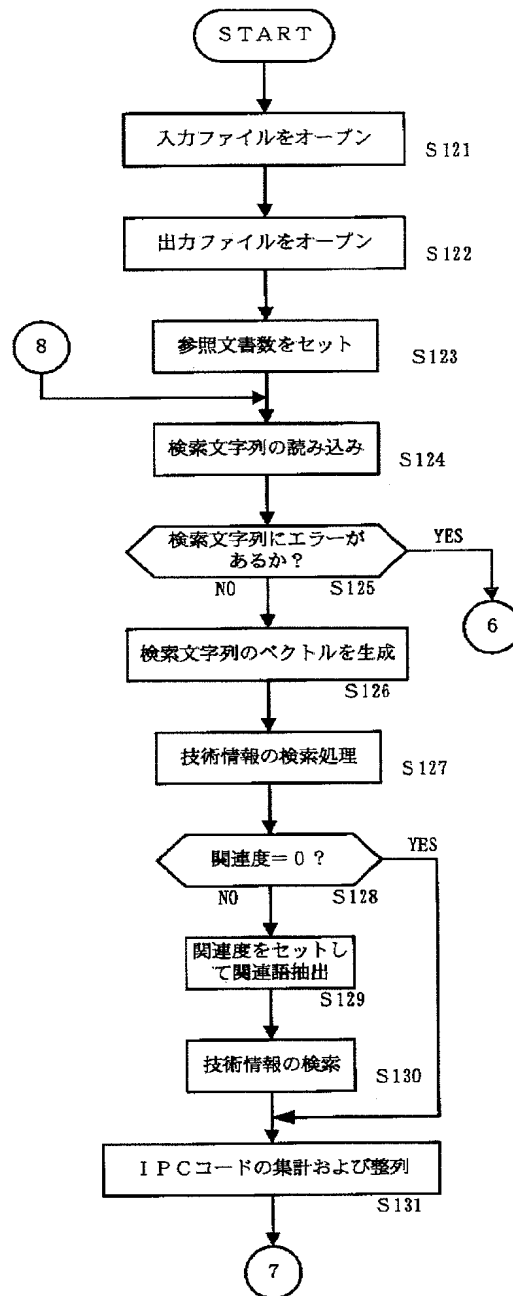
【図15】



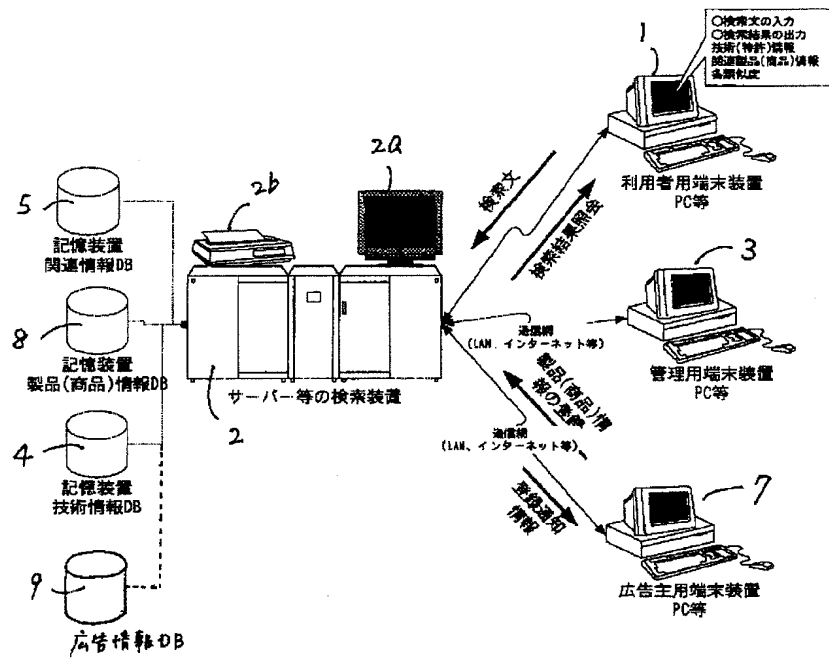
【図17】



【図16】



【図18】



【図20】

検索結果表示TOP画面

検索文: 「ゆがみのないレンズに関する技術が知りたい」

技術名	類似度
非球面レンズ	258
光学レンズ	221
研磨技術	120
...	50

技術情報詳細

技術名: 「非球面レンズ」

概要:

技術保有者:

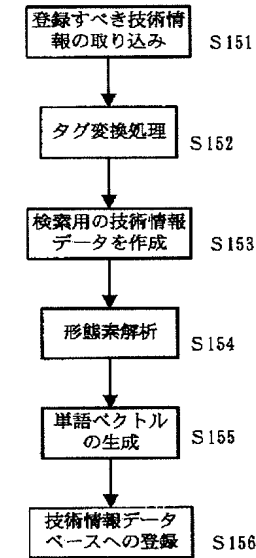
特許出願年:

参考文献:

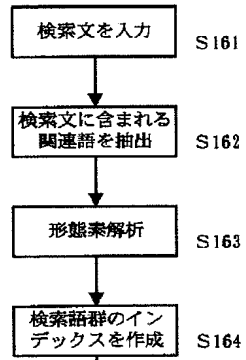
○△顕微鏡 キャッチコピー 関連度	オリンパス・コンパクト カメラLZ123 キャッチコピー 関連度	ソニーDVカメラDV-345 キャッチコピー 関連度
		製品D キャッチコピー 関連度

【図19】

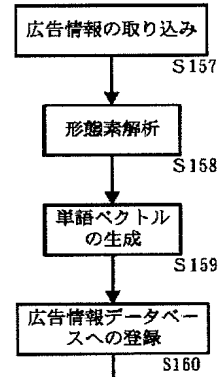
## 技術情報の登録フロー



## 検索フロー



## 広告情報の登録フロー



技術情報データベースに登録されている技術情報、広告情報データベースに登録されている技術情報、および検索語群のインデックスとの間でベクトル処理を行って類似度を検出 S165

算出された類似度が予め定めた基準類似度より小さいか S166

YES

NO

該当する技術情報と広告情報を抽出 S168

検索結果の表示形式を選択 S169

検索文に関連する技術情報がない旨を表示 S167

技術情報別に検索結果を表示 S170

関連技術分類別に検索結果を表示 S171

関連企業別に検索結果を表示 S172

END

【図21】

キャッチコピー：	
製品名：	
会社名：	連絡先：TEL： FAX：URL：E-mail
製品概要：	